Designation: D7372 – 07

An American National Standard

Standard Guide for Analysis and Interpretation of Proficiency Test Program Results[1]

This standard is issued under the fixed designation D7372; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This guide covers the analysis and interpretation of proficiency test (PT) program results. For participants in interlaboratory proficiency test (or crosscheck, check scheme, etc.) programs, this guide describes procedures for assessing participants' results relative to the PT program results and potentially improving the laboratory's testing performance based on the assessment findings and insights (see 6.1). For the committees responsible for the test methods included in interlaboratory proficiency testing programs, this guide describes procedures for assessing the industry's ability to perform test methods, and for potentially identifying needs for test method improvement (see 6.2).

1.2 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory limitations prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D6299 Practice for Applying Statistical Quality Assurance and Control Charting Techniques to Evaluate Analytical Measurement System Performance

D6792 Practice for Quality System in Petroleum Products and Lubricants Testing Laboratories

E177 Practice for Use of the Terms Precision and Bias in ASTM Test Methods

E456 Terminology Relating to Quality and Statistics

3. Terminology

3.1 *Definitions:*

3.1.1 *accuracy*, *n*—closeness of agreement between an observed value and an accepted reference value. **E177, E456**

3.1.2 *assignable cause*, *n*—factor that contributes to variation and that is feasible to detect and identify. **E456**

3.1.3 *bias*, *n*—systematic error that contributes to the difference between a population mean of the measurements or test results and an accepted reference or true value. **E177, E456**

3.1.4 *control limits*, *n*—limits on a control chart that are used as criteria for signaling the need for action or for judging whether a set of data does or does not indicate a state of statistical control. **E456**

3.1.5 *in-statistical-control*, *adj*—process, analytical measurement system, or function that exhibits variations that can only be attributable to common cause. **D6299**

3.1.6 *proficiency testing*, *n*—determination of a laboratory's testing capability by participation in an interlaboratory crosscheck program **D6299**

3.1.7 *Z-score*, *n*—standardized and dimensionless measure of the difference between an individual result in a data set and the arithmetic mean of the dataset, re-expressed in units of standard deviation of the dataset (by dividing the actual difference from the mean by the standard deviation for the data set). **D6299**

3.1.8 *Z′-score*, *n*—measure similar to the *Z-score* except that the PT program standard deviation is replaced with one that takes into account the site precision of the laboratory. Z′ is a valid approach when the laboratory's site precision standard deviation is less than the PT program (that is, these data standard deviation) or stated otherwise when the TPI > 1.

$$Z' = \frac{(X_i - X)}{\sqrt{\left((s')^2 + \left(\frac{s^2_{these\ data}}{n}\right)\right)}}$$

where:

Z' = site precision adjusted *Z-Score*,
X_i = laboratory's result,
X = PT average value,
s' = site precision standard deviation estimate, and
$s_{these\ data}$ = PT Program standard deviation estimate.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *common (chance, random) cause*, *n*—for quality assurance programs, one of generally numerous factors, individually of relatively small importance, that contributes to variation, and that is not feasible to detect or control. **D6299**

[1] This guide is under the jurisdiction of ASTM Committee D02 on Petroleum Products and Lubricants and is the direct responsibility of Subcommittee D02.94 on Coordinating Subcommittee on Quality Assurance and Statistics.

Current edition approved July 1, 2007. Published July 2007. DOI: 10.1520/D7372-07.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

3.2.2 *these test data*, *n*—term used by the ASTM International D02 PT program to identify statistical results calculated from the data submitted by program participants.

3.2.3 *site precision (R')*, *n*—value below which the absolute difference between two individual test results obtained under site precision conditions may be expected to occur with a probability of approximately 0.95 (95 %). It is defined as 2.77 times the standard deviation of results obtained under site precision conditions. **D6299**

3.2.4 *site precision conditions*, *n*—conditions under which test results are obtained by one or more operators in a single site location practicing the same test method on a single measurement system which may comprise multiple instruments, using test specimens taken at random from the same sample of material, over an extended period of time spanning at least a 15-day interval. **D6299**

3.3 *Symbols:*

3.3.1 *I*—individual observation (as in *I*-chart).

3.3.2 *QC*—quality control.

3.3.3 *R'*—site precision.

4. Summary of Guide

4.1 Petroleum and petroleum product samples are regularly analyzed by specified standard test methods as part of a proficiency test program. This guide provides a laboratory with the tools and procedures for evaluating their results from the PT program. Techniques are presented to screen, plot, and interpret test results in accordance with industry-accepted practices.

5. Significance and Use

5.1 This guide can be used to evaluate the performance of a laboratory or group of laboratories participating in an interlaboratory proficiency test (PT) program involving petroleum and petroleum products.

5.2 Data accrued, using the techniques included in this guide, provide the ability to monitor analytical measurement system precision and bias. These data are useful for updating standard test methods, as well as for indicating areas of potential measurement system improvement for action by the laboratory.

5.3 Reference is made in this standard to the ASTM International Interlaboratory Cross-Check Program on Petroleum Products and Lubricants. Program reports containing similarly displayed results and statistical treatments may be available in other PT programs.

6. Procedure

6.1 *Analysis and Interpretation by the Participating Laboratory*—The laboratory should review the results published for each proficiency test program and for each test method or parameter for which the laboratory submitted data. This section covers the evaluations and analyses that the laboratory should consider during their review of proficiency test results.

6.1.1 *Reported versus Submitted Data*—Check to verify that the values ascribed to the laboratory in the Proficiency Test (PT) report agree with the values recorded by the laboratory in its PT records. Verify that the units for the data reported for your laboratory are the same as that requested by the PT program. Report discrepancies to the PT program contacts. Investigate to determine the root cause of the problem.

6.1.2 *Missing Data*—If data and corresponding results are not present when they are clearly expected, then investigate to determine the cause. In some cases it could be an error within the PT program data entry system, or it could be an omission on the part of the laboratory.

6.1.3 *Rejected Data*—Perform an investigation for each instance where laboratory data are rejected by the PT program data treatment process. Attempt to determine the root cause and take corrective actions as needed. Document all such investigations and outcomes. Causes should be shared with the laboratory staff performing the testing. Guidelines on conducting these types of investigations are available in Practice D6299.

6.1.4 *Warnings/Alerts on Data*—The ASTM International D02 PT programs provide comments (that is, Notes 1 to 3 in each Table of Results) that warn participants when their result is:

Note 1—outside 3-sigma range for *these test data*
Note 2—outside 3-sigma range for ASTM reproducibility
Note 3—When the *Z-score* is outside the range -2 to 2

Investigations should also be conducted when any of these warning situations occur. The priority for conducting investigations should be for Note 1 > Note 2 > Note 3. Note 1 indicates that the laboratory is out-of-control with respect to the data set (with the rejected data removed), which is a potentially serious situation with respect to the quality control performance of the corresponding standard test method. A similar argument could also be made for Note 2. Note 3 is a less severe situation, but should be investigated from a continuous improvement standpoint.

NOTE 1—If the user notices that the majority of the laboratories have been cited with a Note 2, then an investigation may not produce any meaningful corrective actions. This occurrence may be the result of the precision statement not accurately reflecting the variability of the test method and should be addressed by the subcommittee responsible for the method. Also, if the Anderson-Darling statistic is >1.3, then the "Note 2" flag may not be valid.

6.1.5 *Z-score*—The *Z-score* calculated for each datum submitted by the laboratory should be reviewed with respect to the following:

6.1.5.1 *Sign and Magnitude of Z-score*—The sign ("+" or "-") of the statistic reflects the relative bias of the individual result versus the mean of the sample group. *Z-score* values falling in the ranges of ±0-1, 1 to 2, 2 to 3, and >3 can be compared to control chart values falling in the ranges between the mean and 1-sigma, 1 to 2-sigma, 2 to 3-sigma, and > 3-sigma. For normally distributed data, there is an expectation that about 68% of the data will lie in the -1 sigma to +1 sigma range, about 95% in the -2 sigma to +2 sigma range, and 99% in the -3 to +3 sigma range. The further a laboratory's *Z-score* is from zero, the greater the relative bias and lower the probability that the data is considered within statistical control. Conduct investigations to determine the cause of any perceived bias as needed.

6.1.5.2 *Trend of Z-scores from Previous Rounds*—Record the *Z-score* values for each test method (parameter) for successive PT program rounds on a control chart to show the trend over time. The lab can use the run rules promulgated in Practice D6299 to evaluate any observed trends. Conduct investigations to determine causes as needed.

6.1.5.3 *Average Z-score*—Calculate the average *Z-score* for a series over a selected time period. The sign and magnitude of this result is an indication of the long-term relative bias. Conduct investigations to determine the cause of any perceived bias as needed.

6.1.6 *Z′-score*—The analysis of any Z′ calculated by the laboratory should be evaluated as described in 6.1.5.3 for the *Z-score*.

6.1.7 *TPI (Industry)*—Consider the *TPI (Industry)* value reported for the data set along with the corresponding *Z-score* for the laboratory's result (reference Guide D6792).

6.1.7.1 *Broad Implications*—Consider the following for interpreting the *TPI (Industry)*:

> 1.2	The performance of the group providing data is probably satisfactory relative to the corresponding ASTM published precision.
0.8 to 1.2	The performance of the group providing data may be marginal. Each laboratory should consider reviewing the test method procedures to identify opportunities for improvement.
< 0.8	The performance of the test method as practiced by the group is not consistent with the ASTM published precision and laboratory method performance. Improvements should be investigated by all laboratories.

6.1.7.2 *Specific Implications Considering TPI (Industry) and Z-score*—A *TPI (Industry)* <0.8 coupled with a *Z-score* >3 (or < -3) implies that the laboratory is likely a significant contributor to the group's poor performance. This situation warrants an investigation to look for potential causes of the apparent bias. When the *TPI (industry)* < 0.8 and the *Z-score* is between 2 and 3 (or -2 and -3), then the laboratory should consider the situation a warning and consider an investigation to find the root cause.

6.1.8 *Precision*—Compare the standard deviation for the PT results versus the site precision value derived from the laboratory's corresponding quality control chart. The expectation is that in most cases the site precision value should be less than the PT program standard deviation. If the laboratory's site precision is greater than the PT standard deviation, then the laboratory should investigate to determine the cause.

6.2 *Analysis and Interpretation by Control Group*—This set covers the analysis and interpretation of proficiency test data by a committee or working group charged with determining the overall implications that the published results have with respect to the corresponding test method or to the working group of participants as a whole. This section covers the evaluations and analyses that the working group should consider during their review.

6.2.1 *TPI and Precision Trends*—Compare precisions obtained over a reasonable number of rounds for a given PT program test method (or parameter). Such data series could be plotted to more clearly show trends. The precision estimates followed may include *TPI (Industry)*, standard deviations, or relative standard deviations (sigma/mean).

6.2.2 *Influence of Uncontrolled Variables on Robust Standard Deviations*—Use auxiliary information or data to create subsets of the PT data set and recalculate precisions and other statistics for each subset. Evaluate these results with the expectation of identifying root causes and potential corrective action steps.

6.2.3 *Normality Evaluations for Historical Sequence*—Plot the PT results using Q-Q Chart and consider the corresponding Anderson-Darling statistic. Observe similar plots for the historical data sets for a given test method (parameter). Investigate situations of non-normal data.

6.2.4 *Contribution of Individual Laboratory Bias to Poor Reproducibility*—Identify the laboratories that are contributing to poor reproducibility (for example, those laboratories with *Z-score* > ±3), and evaluate the factors that may be contributing to this performance. This may involve targeting these laboratories with questionnaires to gather appropriate information. Consultation with test method experts is generally helpful in interpreting results from these investigations.

7. Report

7.1 Laboratories and working groups should document their investigations. In the spirit of continuous improvement, laboratories and working groups are encouraged to share their findings from their investigations and analyses.

8. Keywords

8.1 proficiency testing; quality control; test performance index; Z-score

APPENDIX

(Nonmandatory Information)

X1. CHECKLIST FOR INVESTIGATING THE ROOT CAUSE OF UNSATISFACTORY ANALYTICAL PERFORMANCE

X1.1 To identify why a laboratory's data may have been considered a statistical outlier or to improve the precision, or both, the following action items (not necessarily in the order of preference) are suggested. There may be additional ways to improve the performance.

X1.1.1 Check the results for typos, calculation errors, and transcription errors.

X1.1.2 Reanalyze the sample; compare to site precision, or, if not available, test method repeatability.

X1.1.3 Check the sample for homogeneity, contamination, or that a representative sample has been analyzed.

X1.1.4 Review the test method, and ensure that the latest version of the ASTM test method is being used. Check the procedure step-by-step with the analyst.

X1.1.5 Check the instrument calibration.

X1.1.6 Check the statistical quality control chart to see if the problem developed earlier.

X1.1.7 Check the quality of the reagents and standards used and whether or not they are expired or contaminated.

X1.1.8 Check the equipment for proper operation against the vendor's operating manual.

X1.1.9 Perform maintenance or repairs, or both, on the equipment following guidelines established by the vendor.

X1.1.10 After the problem has been resolved, analyze a certified reference material, if one is available, or the laboratory quality control sample, to ascertain that the analytical operation is under control.

X1.1.11 Provide training to new analysts and, if necessary, refresher training to experienced analysts.

X1.1.12 Document the incident and the learnings for use in the future if a similar problem occurs.